

Acceptable Data Formats

PDS Management Council Meeting
10-11 April 2014

Dick Simpson

Outline

- Proposal to allow 'non-compliant' formats has been forwarded to MC as CCB-46
- Associated issues:
 - PDS Level 1-2-3 Requirements
 - Policy on Acceptable Data Formats (draft)
 - CCB-46: Product_Native + accompanying use policy
- Proposal
- Backup material

Part 1: PDS Level 1-2-3 Requirements

(1 of 2)

(see BACKUP slides for complete set of Requirements)

- *2.4 Peer Review: PDS will conduct **peer reviews of all data submissions** to ensure completeness, accuracy, and scientific usability of content*
 - *2.4.1 PDS will develop and publish procedures for peer review of archival products (which includes **all data submissions and ancillary information**)*
 - ...

Any data submitted to PDS is immediately on track to a peer review for completeness, accuracy, and usability. — RS

PDS Level 1-2-3 Requirements

(2 of 2)

- *3.3 Services: PDS will provide value added services to aid in using archive products.*
 - ...
 - *3.3.2 PDS will provide a capability for opening and inspecting the contents (e.g., label, objects, groups) of any PDS compliant archival product*
 - ...

There will be nothing in the archive that cannot be opened and viewed by a user working with PDS tools — RS

Part 2: **Draft** Policy on Acceptable PDS4 Data Formats (as of October 2011)

PDS shall accept (only?) the following PDS4 data formats:

- i) Fixed-width binary and ASCII tables that are composed of identically structured records;*
- ii) N-dimensional arrays of homogeneous binary elements ($N \leq 16$);*
- iii) Variable-width character 'spreadsheets' that are composed of repeating, M-field, stream-delimited records where the fields themselves are (separately) delimited and may have variable widths ($M > 0$);*
- iv) NAIF/SPICE files.*

PDS shall accept ASCII text and PDF/A formats for PDS4 documentation. PDS shall accept JPEG, GIF, and TIFF images for figures accompanying documents. PDS shall accept any of the approved structures and formats for browse products.

Exceptions to these PDS4 format limitations shall be approved by the PDS Management Council.

These formats appear to be consistent with the PDS Level 1-2-3 Requirements and PDS4 capabilities (although $3 < N < 16$ may still need work). — RS

Part 3: CCB-46

(approve *Product_Native*, associated classes and attributes, and use policy)
(see BACKUP slides for complete change request)

- *PDS will not provide software support other than basic find and retrieve ...*

Violates PDS Requirement 3.3.2 by excusing PDS from providing tools for opening and inspecting products.

Indirectly violates PDS Requirement 2.4 by failing to support proper review.

— RS

CCB-46 Use Policy

(see BACKUP slide for complete proposed policy statement)

- *PDS is not obligated to verify accuracy of documentation which describes Product_Native*
- *PDS is obligated only to store, maintain, and support Search/Retrieval services for Product_Native*
- *PDS is not obligated to provide conversion tools, algorithms, or consultation to Product_Native users.*

Violates Requirements 2.4 and 3.3.2 by excusing PDS from providing the usual support (and thereby inhibiting review) for Product_Native. — RS

- *Holdings may be kept off-line.*

Violates long-standing PDS requirement (since 2008-08-29) that archival holdings be on-line. — RS

CCB-46 Mitigations

- *Each Product_Native must have a companion Product_Observational*
- *Each Product_Native must be accompanied by a bit-level format description (which does not have to be verified).*
- *archive_status may only take on values IN_QUEUE, IN_QUEUE_ACCUMULATING, SAFED, or SUPERSEDED.*
NB: SAFED data are excused from the on-line requirement and are not considered to be "archival" (see BACKUP slide 16)
- *Product_Native submissions will not be "certified".*

Part 4: Proposal

- Revise PDS Requirement 2.4 as follows:
"Peer Review: PDS will conduct peer reviews of {all data submissions} → all submissions of archival data} to ensure completeness, accuracy, and scientific usability of content"
- Revise Draft Policy on Acceptable Formats to: *"PDS shall accept archival data in only the following PDS data formats:"*
Then adopt the policy.
- Approve CCB-46 and the companion use policy

BACKUP

PDS REQUIREMENTS

1. PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and solar system exploration.

1.1 Single Point of Contact: PDS will provide a single point of contact to each mission, program, agency, or individual (*i.e.*, data providers) wishing to submit archival data

- 1.1.1 PDS will assign a lead node for each data provider submitting data to PDS
- 1.1.2 PDS will assign a lead individual, designated by the lead node, who is authorized to negotiate for PDS
- 1.1.3 The PDS lead node will delegate responsibility for subordinate contacts (*e.g.*, instrument teams within a mission) to the appropriate PDS nodes

1.2 Expert Help: PDS will provide expert help in designing archival data sets

- 1.2.1 PDS will provide examples and suggestions on organization of data products, metadata, documentation and software
- 1.2.2 PDS will provide expertise in applying PDS standards
- 1.2.3 PDS will provide expertise to support the design of scientifically useful archival data sets
- 1.2.4 PDS will provide training to support the design of archival data sets for data providers on: PDS standards, tools and services
- 1.2.5 PDS will provide training to develop and maintain staff expertise in data engineering, standards and tools

1.3 Plans and Documents: PDS will assist data providers in developing archive plans, interface documents, validation procedures, and delivery schedules for PDS approval

- 1.3.1 PDS will provide examples of data management and archive plans (including interface documents, procedures, schedules and templates)
- 1.3.2 PDS will determine whether data management and archive plans and relevant interface documents meet PDS requirements
- 1.3.3 PDS will provide criteria for validating archival products
- 1.3.4 PDS will coordinate with the data providers to establish schedules for delivery of archival products to the PDS
- 1.3.5 PDS will coordinate with data providers to establish schedules for public release of archival products

1.4 Archiving Standards: PDS will have archiving standards for planetary science data

- 1.4.1 PDS will define a standard for organizing, formatting, and documenting planetary science data
- 1.4.2 PDS will maintain a dictionary of terms, values, and relationships for standardized description of planetary science data
- 1.4.3 PDS will define a standard grammar for describing planetary science data
- 1.4.4 PDS will establish minimum content requirements for a data set (primary and ancillary data)
- 1.4.5 PDS will, for each mission or other major data provider, produce a list of the minimum components required for archival data

- 1.4.6 PDS will develop, publish and implement a process for managing changes to the archive standards
- 1.4.7 PDS will keep abreast of new developments in archiving standards

1.5 Archiving Tools: PDS will have tools to assist data producers in assembling, validating, and submitting archival products

- 1.5.1 PDS will provide tools to assist data producers in generating PDS compliant products
- 1.5.2 PDS will provide tools to assist data producers in validating products against PDS standards
- 1.5.3 PDS will provide tools to assist data producers in submitting products to the PDS archive
- 1.5.4 PDS will provide documentation for installing, using, and interfacing with each tool

2. PDS will collect suitably organized and well-documented data into archives that are peer reviewed and maintained by members of the scientific community.

2.1 Solicit: PDS will seek complete and comprehensive archives from data providers consistent with interests and resources available.

- 2.1.1 PDS will compare proposed archival submissions against nominal content standards for similar archives and will seek augmentations when the submission is deficient
- 2.1.2 PDS will identify and maintain a list of proposed planetary science data sets to be added to the archive
- 2.1.3 PDS will work with relevant NASA program officials to ensure that products resulting from data analysis programs are submitted to the archive
- 2.1.4 PDS will provide a mechanism for the planetary science community to propose new additions to the archive

2.2 Receive: PDS will receive, acknowledge and track data submissions.

- 2.2.1 PDS will develop and publish the procedures for delivery of data to the PDS
- 2.2.2 PDS will track the status of data deliveries from data providers through the PDS to the deep archive
- 2.2.3 PDS will provide the necessary resources for accepting data deliveries

2.3 Validation: PDS will validate data submissions to ensure compliance with standards.

- 2.3.1 PDS will develop and publish procedures for determining syntactic and semantic compliance with its standards
- 2.3.2 PDS will implement procedures to validate all data submissions to ensure compliance with standards

2.4 Peer Review: PDS will conduct peer reviews of all data submissions to ensure completeness, accuracy, and scientific usability of content.

- 2.4.1 PDS will develop and publish procedures for peer review of archival products (which includes all data submissions and ancillary information)
- 2.4.2 PDS will establish success criteria for peer review of archival products

- 2.4.3 PDS will implement peer reviews, coordinated and conducted by the lead node, to ensure completeness, accuracy and scientific usability of content
- 2.4.4 PDS will publish a summary of the results of each peer review
- 2.4.5 PDS will track the status of each peer review

2.5 Acceptance: PDS will accept or reject submitted data.

- 2.5.1 PDS will develop and publish procedures for accepting archival data
- 2.5.2 PDS will implement procedures for accepting archival data
- 2.5.3 PDS will inform a data provider why a rejected archival product does not meet archiving standards

2.6 Catalog: PDS will maintain a catalog of accepted archival data sets.

- 2.6.1 PDS will develop and publish procedures for cataloging archival data
- 2.6.2 PDS will design and implement a catalog system for managing information about the holdings of the PDS
- 2.6.3 PDS will integrate the catalog with the system for tracking data throughout the PDS

2.7 Storage: PDS will provide appropriate storage for its archive.

- 2.7.1 PDS will develop and publish procedures for storing archival data
- 2.7.2 PDS will maintain appropriate storage for the PDS archive
- 2.7.3 PDS will review its storage capacity and its anticipated storage requirements on a yearly basis
- 2.7.4 PDS will maintain appropriate storage for non-archived data managed by the PDS

2.8 Architecture: PDS will maintain a distributed architecture based on scientific expertise

- 2.8.1 PDS will maintain a distributed archive where holdings are maintained by Discipline Nodes, specializing in subsets of planetary science
- 2.8.2 PDS will maintain a distributed catalog system which describes the holdings of the archive
- 2.8.3 PDS will provide standard protocols for locating, moving, and utilizing data, metadata and computing resources across the distributed archive, among PDS nodes, to and from missions, and to and from the deep archive
- 2.8.4 PDS will work with other space agencies to provide interoperability among planetary science archives
- 2.8.5 PDS will provide an integrated on-line interface that provides information about and links to its data, services, and tools
- 2.8.6 PDS will implement common and discipline-specific services within the distributed architecture
- 2.8.7 The PDS architecture will enable non-PDS developed tools to access PDS holdings and services
- 2.8.8 The PDS architecture will enable computational services on selected archival products

2.9 External Controls: PDS will adhere to applicable federal statutes, NASA policies and Memoranda of Understanding with other organizations.

- 2.9.1 PDS will accept and distribute only those items which are not restricted by the International Traffic in Arms Regulations (ITAR)
 - 2.9.2 PDS will ensure that online interfaces comply with required NASA guidelines
 - 2.9.3 PDS will meet U.S. federal regulations for the preservation and management of data.
 - 2.9.4 PDS will fulfill obligations detailed in any applicable NASA Memorandum of Understanding (MOU)
- 2.10 System Development and Operations: PDS will follow best practices in system and software engineering for developing and operating the system
- 2.10.1 PDS will monitor the system and ensure continuous operation
 - 2.10.2 PDS will identify and adopt technology standards (e.g., hardware and software) for the implementation and operations of the entire PDS system
 - 2.10.3 PDS will ensure that appropriate mechanisms are in place to prevent unauthorized users from compromising the integrity of PDS systems and data

3. PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science.

3.1 Search: PDS will allow and support searches of its archival holdings

- 3.1.1 PDS will provide online interfaces allowing users to search the archive
- 3.1.2 PDS will provide online interfaces for discipline-specific searching
- 3.1.3 PDS will allow products identified within a search to be selected for retrieval

3.2 Retrieval: PDS will facilitate transfers of its data to users

- 3.2.1 PDS will provide online mechanisms allowing users to download portions of the archive
- 3.2.2 PDS will provide a mechanism for offline delivery of portions of the archive to users
- 3.2.3 PDS will provide mechanisms to ensure that data have been transferred intact

3.3 Services: PDS will provide value added services to aid in using archive products.

- 3.3.1 PDS will provide expert help in use of data from the archive
- 3.3.2 PDS will provide a capability for opening and inspecting the contents (*e.g.* label, objects, groups) of any PDS compliant archival product
- 3.3.3 PDS will provide tools for translating archival products between selected formats
- 3.3.4 PDS will provide tools for translating archival products between selected coordinate systems
- 3.3.5 PDS will provide tools for visualizing selected archival products
- 3.3.6 PDS will provide a mechanism for notifying subscribed users when a data set is released or updated

3.3.7 PDS will solicit input from the user community on services desired

4. PDS will ensure the long-term preservation of the data and maintain their usability.

4.1 Long-Term Preservation: PDS will determine requirements for and ensure long-term preservation of the data

- 4.1.1 PDS will define and maintain a set of quality, quantity, and continuity (QQC) requirements for ensuring long term preservation of the archive
- 4.1.2 PDS will develop and implement procedures for periodically ensuring the integrity of the data
- 4.1.3 PDS will develop and implement procedures for periodically refreshing the data by updating the underlying storage technology
- 4.1.4 PDS will develop and implement a disaster recovery plan for the archive
- 4.1.5 PDS will meet U.S. federal regulations for preservation and management of the data through its Memorandum of Understanding (MOU) with the National Space Science Data Center (NSSDC)

4.2 Long-Term Usability: PDS will establish long-term usability requirements and implement procedures for meeting them

- 4.2.1 PDS will define and maintain a set of usability requirements to ensure on-going utility of the data in the archive
- 4.2.2 PDS will develop and implement procedures for periodically monitoring the user community interests and practices and verifying the usability of the products in the archive
- 4.2.3 PDS will monitor the evolution of technology including physical media, storage, and software in an effort to keep the archiving technology decisions relevant within the PDS
- 4.2.4 PDS will provide a mechanism to upgrade products or data sets which do not meet usability requirements (*e.g.*, data sets from old missions)

L1/L2/L3: Approved by PDS Management Council
L1/L2: E-mail vote ending: 6 July 2005
L3: E-mail vote ending: Friday, May 26, 2006
L3 Updates: Approved by MC August 3, 2006
L3 Updates: Approved by MC March 26, 2010

Policy for Online Data Repositories

Each PDS Node will have a primary repository that serves as its source for data distribution to the user community. All archival data holdings will be online in the primary repository by September 2009*.

*"Archival" data have met all PDS acceptance criteria and have been fully ingested in the system; data which have been "safed" by PDS are not in this category.

CCB-46 (page 1 of 2)



</

If you really want to read this, go to <https://pds-jira.jpl.nasa.gov/browse/CCB-46> with your JIRA username and password 17

CCB-46
(page 2 of 2)

insight's SERS SEED format - the instrument's native, compressed format.) Data submitted using this new product will not be considered sufficient to meet a provider's obligation to submit data to the PDS. Product_Native should always be used in addition to Product_Observational in a provider's submission, never in lieu of Product_Observational.

Attachments

| | | |
|---------------------------------------------------------------------------------------------------------------|-------|-------------------|
|  140304_draft_policy_v2.docx | 17 kB | 05/Mar/14 8:05 AM |
|  ccb46_impact_140305.docx | 19 kB | 05/Mar/14 8:05 AM |

Activity


All

Comments


Work Log

History


Activity

 Richard Simpson added a comment - 05/Feb/14 2:25 PM This isn't the normal "up" or "down" vote SCR; it's a request to proceed with prototyping. If CCB choses t


6 older comments

 Steven Hughes added a comment - 05/Mar/14 7:52 AM

Product Native section for the Standards Reference. March 4, 2014

 Steven Hughes added a comment - 05/Mar/14 8:05 AM

Impact questions and answers.


 Todd King added a comment - 06/Mar/14 11:59 AM - edited

The statement:

"PDS or any discipline node may reject Product_Native submissions for any reason, including a general local policy of not wanting to curate them."


in the draft policy is far too permissive and grants too much discretion and veto authority to Nodes. This could and most likely will lead to node-to-node differences as to what can and cannot be archived, even though PDS allows such products.

Basically, if its an allowed product, every node must be willing to accept it as part of a submission.

 Todd King added a comment - 11/Mar/14 9:37 AM

Suggested rephrasing for paragraph 2 of the policy statement:

Consistent with established policies and allowed discretion for all products and collections, PDS or any discipline node may reject Product_Native submissions based on scientific usefulness and merit.

 Ronald Joyner added a comment - 11/Mar/14 10:12 AM

CCB-46: Develop a new PDS4 product for observational data that do not have table or array structures.


ATM: YES
GEO: YES
IMG: YES
IPDA: YES
PPI: YES
RINGS: YES
SBN: NO

PASSES by majority vote.


Here, a "YES" vote means that the policy documents get passed to the MC for their approval. The CCB discussed this at length and calls MC attention to 3 main points:
1) This policy doesn't completely align with the PDS Level 1, 2, 3 Requirements and should be reviewed that this is indeed the direction the MC wants to go regarding Product_Native.
2) Paragraph 2 of the policy includes questionable language – CCB feels this should be scrutinized for the appropriate wording – Suggestion that reasoning for rejection of Product_Native products should be on scientific merit, not solely at the whim of the discipline nodes.
3) Paragraph 7 should be reviewed for alignment with PDS Level 1, 2, 3 Requirements.

People

Assignee:

 Emily Law

Reporter:

 Steven Hughes

Votes:

0 [Vote for this issue](#)

Watchers:

3 [Start watching this issue](#)

Use Policy to Accompany CCB-46

9.Z Constraints on Use of Product_Native

Product_Native will be used only for observational data which (1) are in the original format returned by an experimental system or spacecraft and (2) cannot, in this format, be described by Product_Observational. These data are considered to be optional and supplementary; they do not satisfy any obligations the data provider has to archive with PDS.

PDS or any discipline node may reject Product_Native submissions for any reason, including a general local policy of not wanting to curate them.

Product_Native includes one or more required associations file_area with values File_Area_Native. File_Area_Native includes one or more required associations has_tagged_data_object; only the value Encoded_Native is allowed for this association.

Within the above structure Encoded_Native, a subclass of Encoded_Byte_Stream, includes the required attribute encoding_standard_id which is limited to a list of permissible values. Each permissible value must be approved explicitly by the PDS Management Council. An approved permissible value definition must include identifier(s) for documentation which will allow a user to understand the structure and meaning of the data object contents at the bit level.

Additional information about structure and content must be included in Product_Native.Reference_List including LIDs and/or LIDVIDs to the documentation referenced in the encoding_standard_id.

For each Product_Native a Product_Native.Reference_List.Internal_Reference is required that references an associated Product_Observational using a LID or LIDVID and that has a reference_type value of “native_to_archival”. The associated Product_Observational must have scientifically equivalent content.

PDS will verify the accuracy of the encoding_standard_id and Reference_List information at the time of peer review but is not obligated to confirm the accuracy of the documentation itself.

PDS is obligated only to store and maintain (*e.g.*, periodic checksum validations) Product_Native submissions and to support Search and Retrieval services. Search may be limited by the granularity of label information; these holdings are not required to be on-line, so Retrieval may be slow. PDS is not obligated to provide conversion tools, algorithms, or consultation to users interested in working with Product_Native submissions.

The tracking attribute archive_status will be defined for all PDS4 holdings; its values will be the same as for its use within the PDS3 migration context. For Product_Native, the only allowed values are IN_QUEUE, IN_QUEUE_ACCUMULATING, SAFED, and SUPERSEDED.

No Product_Native will ever be given the status “certified” under the PDS policy adopted on 2010-09-21, as amended.